

LATVIJAS UNIVERSITĀTE

BAKALaura DARBS

RĪGA 2015

UNIVERSITY OF LATVIA
FACULTY OF HUMANITIES
DEPARTMENT OF ENGLISH STUDIES

**ORTHOGRAPHIC ERROR FREQUENCY IN
STUDENTS' WRITTEN BUSINESS ENGLISH**

**ORTOGRĀFISKO KĻŪDU BIEŽUMS IZGLĪTOJAMO
LIETIŠKAJĀ ANĢĻU RAKSTU VALODĀ**

BACHELOR THESIS

Arina Mihailova

Matriculation card No. 11109

Adviser: Assist. Prof. Natalja Cigankova

RĪGA 2015

ACKNOWLEDGEMENTS

I would like to express gratitude to assist. prof. Natalja Cigankova for her guidance, relevant comments and suggestions in the process of writing the current BA Thesis.

ANOTĀCIJA

Darba *Ortogrāfisko kļūdu biežums izglītojamo lietišķajā angļu rakstu valodā mērķis* ir analizēt ortogrāfiskās kļūdas studentu esejās lietišķajā angļu rakstu valodā un izpētīt, kurš ortogrāfisko kļūdu veids ir visbiežāk sastopamais. Pētnieciskā darba pamatdoma ir izpētīt, kādas ortogrāfiskās kļūdas pieļauj Latvijas Universitātes studenti, kuri apgūst otru svešvalodu.

Darba autore izveidoja korpusu no 59 studentu esejām, kuras bija uzrakstītas lekcijas laikā, un veica kļūdu analīzi, kurā apskata ortogrāfiskās kļūdas un to veidus.

Kļūdu analīze apliecina, ka studentu esejas lietišķajā angļu rakstu valodā satur ievērojamu ortogrāfisko kļūdu daudzumu, neskatoties uz to, ka studentiem ir atļauts izmantot pareizrakstības pārbaudītāju. Visbiežāk sastopamās ortogrāfiskās kļūdas ir pareizrakstības kļūdas un nepareizs komatu lietojums.

Atslēgas vārdi: lingvistiskais korpus, korpus, ortogrāfiskās kļūdas, kļūdu analīze

ABSTRACT

The aim of the research paper *Orthographic Error Frequency in Students' Written Business English* is to analyse orthographic errors in students' essays written in the classroom in Business English and to find out which types of orthographic errors are the most common. The main idea of the current paper is to investigate the orthographic errors that are committed by second language learners of the University of Latvia.

The author of the current paper has compiled a corpus of 59 students' essays written in the classroom and has done an error analysis which describes the orthographical errors and their types.

The results of the error analysis show that there is a number of orthographic errors in students' essays. Even though students had been able to use spell checkers, the most common orthographic errors are spelling and punctuation errors.

Key words: corpus linguistics, corpus, orthographic errors, error analysis

CONTENTS

LIST OF ABBREVIATIONS AND ACRONYMS.....	1
INTRODUCTION.....	2
1. THEORETICAL BACKGROUND TO ERROR ANALYSIS AND CORPUS-BASED APPROACH.....	5
1.1 Error analysis approach.....	5
1.2 Types of errors.....	9
1.3 Orthographic errors.....	14
1.3.1 Spelling errors.....	16
1.3.2 Hyphenation errors.....	17
1.3.3 Capitalization errors.....	17
1.3.4 Punctuation errors.....	18
1.4 Interpretation of error analysis.....	19
1.5 Corpus-based approach error analysis.....	21
1.5.1 Defining corpus.....	24
1.5.2 Types of corpora.....	26
2. RESEARCH METHODOLOGY.....	29
2.1 Research method.....	29
2.2 Research object.....	30
2.3 Research procedure.....	30
3. RESULTS AND DISCUSSION.....	32
3.1 Research results.....	32
3.2 Discussion on findings.....	37
CONCLUSIONS.....	38
THESES.....	40
REFERENCES.....	41

LIST OF TABLES

<i>Table 1.1</i> Error Analysis Model of ESL learners (adapted from Ferris, 2005).....	13
<i>Table 1.2</i> Error categories and explanation of them (adapted from Ferris, 2005)	14
<i>Table 3.1</i> Number of orthographical errors in students' essays	32
<i>Table 3.2</i> Examples of spelling errors in students' essays.....	33
<i>Table 3.3</i> Examples of capitalization errors in students' essays.....	35
<i>Table 3.4</i> Examples of hyphenation errors in students' essays.....	36

LIST OF ABBREVIATIONS AND ACRONYMS

BA - Bachelor

BNC - British National Corpus

ANC - American National Corpus

L1 - First Language (mother tongue)

SL or L2 - Second Language

TL - target language

NL - native language

EA - error analysis

INTRODUCTION

The current Bachelor thesis is focused on compiling a corpus of students' essays written in the classroom of Business English. The compiling of the essay corpus will help to analyse the frequency of orthographic errors in students' essays easier and more accurately, as the corpus is based on real data provided by the students of the University of Latvia. The topic of analysing orthographic errors in classroom-written essays in English is relevant because, as for the students, it is their second language and it is difficult to avoid these errors. There are many possible reasons that will be discussed in the current BA thesis.

According to Johanna Stirling (2011), it is considered that English orthography is very difficult not only for people for whom English is their second language but for native speakers as well. The reason of incorrect orthography may be the fact that in the English language almost each letter has at least two ways of pronouncing it, including silence as a result of omitting the letter when speaking (Stirling, 2011:41-42).

Orthography plays a key role because many misunderstandings can be caused when using it incorrectly. Due to the development of technology, homework essays for school or university studies are written on computers and with the help of automatic spelling and grammar checkers. For that reason, students usually do not even know (by heart) how to spell some words correctly.

As this paper aims to analyse the orthography of students' essays written in the classroom, it is relevant to use a corpus for such investigation. The corpus will allow for the analysis in the empirical part of the present research. There are large corpora already compiled and available to the public, such as British National Corpus, American National Corpus and many others that contain texts from different journals, books, letters, essays, as well as different oral data (Baker, 2007:28). However, it is also possible to use corpora of less significant size to carry out an analysis; for example, Shalom (1997) created a corpus of small samples to show the frequency of such words as 'slim' and 'attractive.' She collected 776 advertisements that were sent to a certain magazine (ibid.).

Building the corpus of student's essays will help to analyse the frequency of errors more precisely and will help to understand the most common types of orthographical errors.

The research of the current topic will create the opportunity to use the data for further research to develop a syllabus to avoid or at least decrease the number of orthographical errors committed by second language learners in the future.

The research goal: To investigate the most common orthographic errors in students' essays with the help of a corpus compiled of Business English essays written by the students of the University of Latvia, as well as to determine what type of orthographical errors are most common.

The research objectives:

- To read and analyse the theoretical material on interlanguage and error analysis in order to create a theoretical framework for the empirical analysis where the orthographical errors in the students' essays corpus will be analysed;
- To compile a corpus of essays written in Business English by the students of the University of Latvia, Modern Language and Business Studies study programme;
- To analyse the frequency of orthographical errors in students' essays in the compiled corpus;
- To analyse the most frequent types of orthographic errors;
- To draw relevant conclusions on the orthographic errors found in the corpus, as well as their possible reasons.

The research questions:

1. What are the most frequent orthographical errors?
2. What type of orthographic errors is most common in essays written by the students of the University of Latvia?

Hypothesis: Orthographic errors are less frequent than other types of errors made in students' essays, written under supervision in a classroom equipped with word processors, as modern technology offers the means to find and correct them.

The research method: The method of the current BA paper includes a qualitative error analysis and a quantitative measurement for determining the frequency of occurring errors. The qualitative method involved data collection for a corpus, corpus compiling, analysis of orthographic error frequency and analysis of types of orthographic errors visible in the corpus of essays by the students of the University of Latvia.

Outline of the paper

The first chapter of the BA paper presents a literature review on two relevant research approaches, such as the error analysis approach and the corpus-based approach. This chapter presents a brief overview of error types, describes orthographical errors and classifies them into types, as well as presents several types of corpora and discusses the interpretation of error causes. This chapter is devoted to facilitating the empirical analysis of the current research paper.

The second chapter comprises the methodology of the research. It provides the research method used for the current research, describes the corpus and the procedure of the research.

The third chapter of the paper provides the research findings and a discussion on them. It describes the results of the empirical part of this paper and shows relevant examples of real orthographic errors. That, in turn, helps to draw relevant conclusions on the most common orthographic errors in students' essays.

1 THEORETICAL BACKGROUND TO ERROR ANALYSIS AND CORPUS-BASED APPROACH

The first chapter of the current BA paper contains the analysis on theoretical materials regarding two approaches of error analysis. The first part of the chapter gives an overview of the error analysis approach and shows different types of errors, specifies the types of orthographic errors and describes typical errors made by second language (L2) learners. The second part of the chapter presents a corpus-based error analysis approach, as well as defines the corpus and shows how a corpus can facilitate the study of the second language.

The first chapter combines the materials by different researchers, such as Selinker (1972), Tarone (1997), Flowerdew (2012), Tuabert and Chermakova (2004) and others.

1.1 Error analysis approach

In the language learning process, it is unavoidable to commit errors. Some time ago, teachers were very strict with the errors and tried to push students to avoid making errors (Touchie, 1986:75). However, in the past years researchers have developed Applied Linguistics, where errors are regarded as a part of the second language learning process during which students apply different strategies in L2 learning (ibid.).

Corder (1976) has an explanation to the concept of errors that are committed by students, and there are two schools of thought (Corder, 1976; Chomsky, 1966). The first school of thought assures that errors occur due to a wrong teaching technique chosen by the teacher, and, if the right technique is used, no errors will occur. However, the second school of thought claims that there will always be errors, and, according to this school, there will always be the necessity for error analysis. According to Corder, contrastive analysis is based on the behaviour and habits of the learners, but error analysis is based on theories of the processes in the brain when people are learning.

Corder (1974) states that learners of second language develop rules in their minds; sometimes a learner might say or write something correctly even without knowing and only because the learner has heard the right form or expression somewhere (Corder, 1974:22). Corder says: 'It is by reducing the language to a simpler system than it is that the child reveals his tendency to induce rules' (ibid.:23). Corder classifies errors in two groups: errors that occur systematically and errors that do not occur regularly. Systematic errors are 'underlying knowledge of the language to date, or, as we may call it [the learner's] transitional

competence' (Corder, 1974:25), and the errors that do not underline this competence are performance errors that can be classified as mistakes (ibid.). However, Corder says that systematic errors or the errors that underline the transitional competence are classified as errors. He states that 'A learner's errors provide evidence of the system of the language that he is using (i.e. has learned) at a particular point in the course (and it must be repeated that he is using some system, although it is not yet the right system)' (ibid.). He also mentions that the knowledge of a learner can be restored (ibid.).

According to Corder, there are three groups for whom errors are useful (ibid.). Firstly, errors are useful for teachers to understand what level of knowledge learners have and what should be improved (ibid.). Secondly, they are useful for a researcher in order to investigate language acquisition. Finally, they are useful for learners to analyse their errors and not commit them in the future (ibid.).

Corder claims that interference errors are not concerned with the habits of the learner in errors but in 'investigating the systems of the new language' (Corder, 1974:27). He states that teachers of second language have to discuss errors and have to make learners analyse them in order to understand the language structure (ibid.).

According to Khansir, Error Analysis is a very important topic in second language learning because it is considered as 'essential approach of language learning' (Khansir, 2012:1027). Second language learners do not have knowledge of the fact that language contains systems or rules (ibid.). To analyse the errors, there are two approaches: contrastive and error analysis (ibid.). It is important to admit that in error analysis it is impossible to omit contrastive analysis. These two analyses are accepted in the science of Applied Linguistics (ibid.).

According to Fisiak (1981), Contrastive Linguistics is 'a subdiscipline of Linguistics concerned with the comparison of two or more language or subsystems of language in order to determine both differences and similarities between them' (as quoted in Khansir, 2012:1027). Carl assured that Contrastive Analysis is an inevitable constituent of the second language learning process. Contrastive Analysis predicts that 'the speaker of an arbitrary first language is liable to produce grammatically deviant second language sentences, the structural descriptions of which will resemble those of analogous first language sentences' (ibid.).

Error Analysis is considered to be a branch of Applied Linguistics (ibid.). It appeared to show that language learners' errors are part of universal language acquisition strategies and they happen not only because of the mother tongue of the learner (ibid.). Khansir says that this is 'a reaction to Contrastive Analysis Theory' and it is believed that 'native language

interference as the major source of errors in second language learning what behaviouristic theory suggested' (Khansir, 2012:1027).

According to Erdogan (2005), 'applied error analysis, on the other hand, concerns organizing remedial courses and devising appropriate materials and teaching strategies based on the findings of theoretical error analysis' (as quoted in *ibid.*). Richards (1971) explains that the error analysis field deals with the differences between how learners use the language and how they learn it (as quoted in *ibid.*). Norrish (1983) has said that it is called systemic deviation and happens 'when a learner has not learned something and consistently gets it wrong, an error' (as quoted in Khansir, 2012:1029). Norrish gives an example: he mentions the use of the infinitive with *to* after *must* (e.g., I must to buy food). Basically, if the learner knows that *to* is used after such verbs as *want* and *need*, the learner also supposes that *to* is also used after *must* and produces it in this way, until someone tells him/her or he/she notices it when native speakers do not use it in this way and he/she will produce it systematically (*ibid.*).

Since there are problems in second language acquisition, the studies have focused on the errors provided by learners. Error analysis is considered to be linguistic analysis that concentrates on the errors produced by learners (*ibid.*:1029). Khansir explains that error analysis accents the importance of errors made by L2 learners and also in the target language itself (*ibid.*). He points out that 'interferences from the learners mother tongue is not only reason for committing errors in his target language' (*ibid.*).

Sharma (1980) has said that error analysis can provide a strong support to 'remedial teaching.' He says that it can affect the teaching program in positive and negative ways because it is also possible to reveal the problems in the program (as quoted in Khansir, 2012:1029).

Dulay (1982) says that error analysis is helpful in the process of leaning second language as it can be used to design exercises and pay more attention to the weakness of the learner to improve them (as discussed in Khansir, 2012:1029).

According to Strevens (1969) hypothesis, errors do not have to be presented as problems to get over them. He explains that errors have to be viewed as normal features 'indicating the strategies that learners use' (as quoted in Richards, 1975:4).

Richards has investigated that there are seven factors that affect the learning systems of L2 students. Those factors are:

- language transfer,
- intralingual interference,
- sociolinguistic situation,
- modality, age,

- successions of approximatively systems,
- universal hierarchy of difficulty. (Richards, 1975:4)

Richards named the reasons why the study of errors is useful, namely to spot language learning strategies, learners' errors and hardship in acquiring a new language. Error analysis in second language learning helps to recognize the linguistic difficulties and 'needs at a particular stage of language learning' (Khansir, 2012:1027). This analysis has implications to deal with learners' errors in the classroom. The first one is to devise remedial measures, the following one is to prepare materials that start with the easier tasks, followed by more difficult ones, and the last implication is to make suggestions for learners' strategies (ibid.).

The analysis of errors, developed by Corder (1947), has five steps and these steps are:

1. Collection of a sample of learner language
2. Identification of errors
3. Description of errors
4. Frequency of errors
4. Explanation of errors
5. Evaluation of errors. (Gass, Selinker, 2008:103)

It is relevant to explain these steps. The first step of Error Analysis (EA) provides the compiling of a corpus by the collection of written data or sometimes oral data (Gass, Selinker, 2008:103). According to Ellis and Barkhuizen (2005), 'the nature of the sample that is collected may influence the nature and distribution of the errors observed' (Ellis, Barkhuizen, 2005:57). The second step deals with the identification of errors; e.g., incorrect use of a singular verb with the subject in plural or wrong sequence of tenses (Gass, Selinker, 2008:103). Ellis and Barkhuizen state that it is important to specify the place in the sentence where the error occurs and evaluate how serious the error is (Ellis, Barkhuizen, 2005:57). The third step deals with the classification of the errors. Gass and Selinker say that it is important to identify the type and give an example by asking the question 'is it an error of agreement?' or 'is it an error in irregular verbs?' (Gass, Selinker, 2008:103).

Next, the fourth step is related to error frequency. Gass and Selinker say that it is important to count how many errors of incorrect irregular verb use or wrong agreement between words the text has (ibid.). Ellis and Barkhuizen say that this step is 'recording of frequency of the errors in each category' and 'the development of a set of descriptive categories for coding the errors that have been identified' (Ellis, Barkhuizen, 2005:60).

According to Ellis and Barkhuizen, the fifth step - explaining errors - involves determining their sources in order to find out why they were made (ibid.:62).

Lastly, the sixth step is devoted to the correction of errors. Gass and Selinker claim that, depending on the type and frequency of errors, a pedagogical intervention is made (Gass and Selinker, 2008:103).

1.2 Types of errors

According to Touchie (1986), there are two types of errors in the field of Applied Linguistics: performance and competence errors (1986:76). He explains that performance errors are the errors that occur due to inattentiveness and these errors are easy to get over; however, competence errors are the errors that occur due to inappropriate learning of the second language and these errors are considered as more serious (ibid.). Touchie also mentions that Gefen (1979) has differentiated between errors and mistakes (as quoted in ibid.). Therefore, it is important to find out the difference between *error* and *mistake*.

James (1998) says that mistakes occur when the learner can correct his/her fault by himself/herself and it means that this fault was not intended (James, 1998:78). However, the indented errors that cannot be corrected by the learner are errors (ibid.). It is relevant to mention that L2 learners mostly produce errors, which do not show the learners' knowledge and cannot be corrected by the learners (ibid.).

Burt and Kiparsky (1974) divide errors in two groups: global and local errors (as quoted in James, 1998:78). They say that local errors do not affect the meaning of the sentence, whereas global errors cause misunderstandings because they may change the meaning of the sentence. According to Burt and Kiparsky, local errors include 'noun and verb inflections, and the use of articles, prepositions, and auxiliaries' (James, 1998:78). Global errors include wrong word order (ibid.).

Second language learning errors include 'the phonological, the morphological, the lexical, and the syntactic' components (ibid.:77). Touchie (1986) gives an example of a phonological error among learners from Arab countries and it is 'the hick of distinction between the phoneme /p/ and the phon-eme /b/' (Touchie, 1986:77). For that reason, they say '*pird* and *brison*, for example, instead of *bird* and *prison*' (ibid.). Touchie provides an example of morphological errors as well as 'the production of such errors as *womans*, *sheeps*, and *furnitures*' (ibid.). Touchie says that lexical errors occur due to the translation from the mother tongue or 'the use of wrong lexical items in the second language' (ibid.). He gives such examples as 'This is the *home* that my father built, and The *clock* is now ten' (ibid.). According to Touchie, syntactic errors are 'errors in word order, subject-verb agreement, and the use of the presumptive pronoun' (ibid.).

There exists another way of distinguishing between errors. According to Brown (1994) and Connor (1996), there are two different categories of errors. The first category deals with the errors that are committed in the stages when there is an intervention of the first language and they are called interlingual errors. According to Brown, this type of errors usually occurs in the first stages of second language learning because of the familiarity with the first language. To avoid these errors, learners have to become familiar with the system of the second language. In that case, these errors will disappear. The above mentioned result is called a positive transfer; however, a negative transfer (interference) is also possible (Brown, 1994:224).

The second error category is intralingual and developmental errors (Richards, 1975:190) that, according to Jain (1974), occur in the result of such conditions as ‘learning strategies, different training procedures, and individual differences of teachers, learners, text books’ (as quoted in Richards, 1975:190). Richards defines interlingual errors as errors ‘which reflect the general characteristics of rule learning, such as faulty generalization, incomplete application of rules and failure to learn conditions for rule application, the learner attempting to build up hypothesis about English from his limited experience of it in the classroom or textbook’ (ibid: 198), and he says that these errors are caused due to a misunderstanding of the rules of the language (ibid.).

According to Richards, developmental errors are errors that ‘do not derive from transfer from another language’ (ibid.:199) and ‘they reflect the learner’s competence at a particular stage and illustrate some of the general characteristics of language acquisition errors committed during various stages of L2 learning process’ (ibid.). Second language learners commit these errors in different stages but the main cause is a wrong assumption of the target language (ibid.).

According to Touchie (1986) ‘intralingual and developmental errors are due to the difficulty of the second/target language’ (1986:77) and these errors involve eight factors:

1. Simplification
2. Overgeneralization
3. Hypercorrection
4. Faulty teaching
5. Fossilization
6. Avoidance
7. Inadequate learning
8. False concepts hypothesized. (Touchie 1986:77)

Touchie explains simplifications as the use of simple forms in alternative to more complex ones (1986:77). He presents an example of when students use ‘simple present instead of present perfect continuous’ (ibid.). Overgeneralization is a wrong use of

constructions in the context where in one case it is possible and in the other - inappropriate (ibid.). As an example Touchie mentions the avoidance of *-s* in the third person singular or the use of *-ed* in words where it is not correct, such as *goes* instead of *went* (ibid.). Touchie admits that ‘simplification and overgeneralization are used by learners in order to reduce their linguistic burden’ (1986:77). What is more, there are cases when teachers are devoted to much in the students’ error correction and it affects the students to ‘make errors in otherwise correct forms’ (ibid.); that is called hypercorrection (ibid.). Stenson (1978) named these errors *induced errors* (as quoted in Touchie, 1986:77). Touchie gives examples to such errors in the case of language learning process for Arabs when a teacher stresses to ‘produce the phoneme *IpI* correctly [and it] prompts [the students] to always produce *IpI* where the phoneme *Ibl* is required’ (ibid.), and these learners produce ‘pird and pattle instead of bird and battle’ (ibid.).

Next, there are errors that occur due to ‘the teacher, teaching materials, or the order of presentation’ (ibid.). According to Touchie, hypercorrection and faulty teaching are connected (ibid.). He admits that, in a long period of teaching, the teacher may be affected by the students’ errors (ibid.).

According to Touchie, fossilization errors are errors that are ‘difficult to get rid of’ (ibid.). These are mainly pronunciation errors; for example, the same /p/ and /b/ pronunciation in English by Arab learners (ibid.).

Touchie continues the discussion, accentuating that there are learners that simply avoid some difficult syntactic structures; that is called avoidance (1986:78). For example ‘Arab ESL learners avoid the passive voice while Japanese learners avoid relativization in English’ (ibid.).

What is more, inadequate learning results in defective learning or ignorance of rules in the language learning process (ibid.). For example, ‘omission of the third person singular *-s* as in: He *want*’ (ibid.).

Some learners have false concepts hypothesized about the Target Language (TL); for instance, when ‘learners think that *is* is the marker of the present tense’ and they say ‘He *is talk to the teacher*’ (ibid.).

Next, the two taxonomy error types are investigated. The first type is the allocation of the linguistic category. This type indicates on ‘what *level* of language the error is located’ (James, 1998:104). For example, is it located in ‘phonology, graphology, grammar, lexis, text or discourse?’ (ibid.:105). The following issue is detailing the error occurrence and determining what the category of its linguistic unit is. James claims that it is important to specify what kind of grammatical construction the error involves if the error is made on a grammatical level; for example, ‘the auxiliary system, passives, sentence components’ (ibid.).

After finding the level of the error, the establishment of the *class* follows (ibid.). If it is a, for example, grammatical error, ‘does it involve the class of noun, verb, adjective, adverb, preposition, conjunction, determiner, etc.?’ (ibid.). Afterwards, it is important to determine the *rank* of the error. Basically, it is based on the ‘hierarchy of *units*’ where the error lies (ibid.). And the final thing is specifying the *system* that becomes affected by the error; for example, ‘tense, number, voice, countability, transitivity, etc.’ (ibid.).

The other type of taxonomy is surface structure that was proposed by Dulay and Krashen (1982) They say that this taxonomy is based on ‘the ways surface structures are altered’ (as quoted in 1998:106). Dulay and Krashen discover four ways ‘in which IL and TL forms diverge in specific and systematic ways;’ however, James added one more type (ibid.). They are:

1. *Omission*

Omission is considered as ungrammatical and, according to Dulay, Burt and Krashen (1982), is disturbing (as quoted in James, 1998:106); for example, in ‘*He’ll pass his exam and I’ll,*’ too is omitted (ibid.). At least in the early stages of L2 acquisition, omission is influencing ‘function words rather than content words’ (ibid.:107). According to Kasper and Kellerman (1997), advanced learners use compensatory strategies to state their ideas (as quoted in James, 1998:107). It is explained by the awareness of ‘their ignorance of content words’ (ibid.). Instead of ignoring them, the learners resort to this strategy (ibid.).

2. *Addition*

According to Dulay, Burt and Krashen, addition is the ‘result of all-too-faithful use of certain rules’ (as quoted in James, 1998:107). They divide it in three subtypes. The first subtype is regularization (e.g., ‘producing the regular **bued for bought*’) and irregularization (e.g., ‘**drove* for the preterite form *dived*’) (ibid.). The second subtype is double marking which is defined as ‘a failure to delete certain items’ (e.g., ‘*He doesn’t know*s me*’ or ‘**I didn’t went there yesterday*’) (ibid.). The third type is simple addition. It deals with all other additions (ibid.:108).

3. *Misformation*

This type is defined as ‘the use of wrong form of a structure or morpheme’ (James, 1998:108). Dulay, Burt and Krashen give an example of misformation: ‘*I*seen her yesterday*’ (ibid.) and it is obvious that *seen* is used instead of *saw*. Misformation is divided in three subtypes: archiform, e.g., ‘out of the set *this/that/those/these* the learner might use only *that*’ (ibid.), regularization, e.g., ‘*womans*’ (ibid.) and alternating forms, for example, ‘*I no play*’ (ibid.:109).

4. *Misordering*

This type is observed when language learners ‘word-for-word’ translate from SL to TL (ibid.). Ellis and Barkhuizen (2005) say that these are ‘errors characterised by the incorrect placement of a morpheme or group of morphemes in an utterance’ (Ellis, Barkhuizen, 2005:62), and mention an example: ‘She fights all the time her brother’ (ibid.).

The error taxonomy provided by Dulay, Burt and Krashen (1982) is highly useful and James (1998) calls it *Target Modification Taxonomy*. It is based on ‘a comparison of forms the learner used with the forms that a native speaker would have used in the same situation’ (James, 1998:111). Supplementary to the four-category framework, James adds the fifth category:

5. *Blends*

Blend errors occur when a learner has two targets and he/she has not decided what target he/she ‘has in mind’ (James, 1998:111). Ellis and Barkhuizen also say that these errors ‘reflect the learner’s uncertainty as to which of two forms is required’ (Ellis, Barkhuizen, 2005:62). James gives an example: ‘**according to Erica’s opinion.*’ This example shows that the learner has blended two sentences ‘*in Erica’s opinion*’ and ‘*according to Erica*’ (James, 1998:111).

Ferris (2005) has developed the Analysis Model that presents the most common errors in ESL. It is categorized in four types and they are as follows: morphological, lexical, syntactic and mechanical errors. The model is presented below in *Table 1.1*. However, *Table 2.2* describes major error categories that include verb errors, noun ending errors, article errors, word wrong, and sentence structure, also presented by Ferris (2005:92).

Table 1.1 Error Analysis Model of ESL learners (adapted from Ferris, 2005)

Morphological Errors → global / local errors Verbs: Tense, Form, Subject-verb agreement Nouns: Articles/determiners, Noun endings (plural/possessive)
Lexical Errors → local errors Word choice, Word form, Informal usage, Idiom error, Pronoun error
Syntactic Errors → global errors Sentence structure, Run-ons, Fragments
Mechanical → local errors Punctuation, Spelling

Table 1.2 Error categories and their explanation (adapted from Ferris, 2005)

Verb errors	All errors in verb tense or form, including relevant subject-verb agreement errors.
Noun ending errors	Plural or progressive ending incorrect, omitted, or unnecessary; includes relevant subject-verb agreement errors.
Article errors	Article or other determiner incorrect, omitted, or unnecessary.
Wrong word	All specific lexical errors in word choice or word form, including preposition and pronoun errors. Spelling errors only included if the (apparent) misspelling resulted in an actual English word.
Sentence structure	Errors in sentence/clause boundaries (run-ons, fragments, comma splices), word order, omitted words or phrases, unnecessary words or phrases; other unidiomatic sentence construction.

Ferris' (2005) model, presented in the tables above, divided errors in two categories: local or global errors. According to Burt and Kiparsky (1972) global errors are errors that are major and, if they occur, it is sometimes difficult to understand the sentence; however, local errors are minor errors and they do not affect the meaning of the sentence. 'With this system it is easy to identify global and local errors' (Burt and Kiparsky, 1972, as cited in James, 1998:78).

1.3 Orthographic errors

According to Brigitte van Berkelt and Koenraad De Smedt (1988), orthographic errors are 'cognitive errors consisting of the substitution of a deviant spelling for a correct one' when the creator does not know the right spelling of the word, has forgot the right spelling or simply misunderstood the word (Berkelt, Smedt, 1988:77). Berkelt and Smedt characterize orthographic errors as follows: 'they generally result in a string which is phonologically identical or very similar to correct string'; for example, 'indecies' instead of 'indices' (ibid.). It leads to the assumption that in particular languages, orthographical errors 'are dependent on the correspondence between spelling and pronunciation' (ibid.). According to Berkelt and Smedt, another characteristic for orthographic errors is 'that proper names, infrequent words and foreign words are particularly prone to orthographic errors' (ibid.).

According to ThambiJose (2014), there are several spelling type characteristics and it is why orthographic errors are committed. The first reason is because of mispronunciation (ThambiJose, 2014:2439). According to Roy (1998), ‘knowing the correct pronunciation of a word and the graphemes of an alphabet would be intuitive and easy to infer the correct spelling of any word;’ however, sometimes in the English language knowing the right pronunciation of the word does not help to spell the word in the right way (ibid.:2440). The English language has ‘large conventional orthography’ and this is the cause why students make orthographical mistakes when writing in this language (ibid.) According to ThambiJose, ‘misspelling of words in compositions are usually committed without any reason’ (ibid.) and it happens because the author does simply not know the correct spelling (ibid.). In fact, orthographic errors are not committed intentionally (ibid.). For example, the word ‘please’ as ‘pleace’ or ‘figure’ as ‘figer’ (ibid.:2441). ThambiJose also admits that ‘phonological features such as addition, deletion and vowel change are found in the words with orthographic errors’ (ibid.).

The next reason why errors are committed is ‘because of uncommon words’ (ibid.). TambiJose explains that the English language has ‘rich written uses, technical words and borrowed words’ that are often used in writing and rarely in speaking (ibid.). ThambiJose gives two examples when ‘prompts’ is spelled as ‘proms’ or ‘distortion’ is spelled as ‘distosion’ (ibid.) In these cases, errors are made ‘from simple ignorance of the exact spelling of words without implications regarding the relation between sound and letters’ (ibid.).

One more reason why words are misspelled is borrowed words. There are many words in the English language that are borrowed from, for example, Greek, Latin or French. In that case, non-native authors make orthographic mistakes (ibid.). For example, word ‘datum’ is spelled as ‘datem’ or ‘myth’ is spelled as ‘mith’ (ibid.:2442).

The position of phoneme may also cause orthographical errors, and these positions are initial, medial and final (ibid.).

An example of misspelled words in the initial position is ‘thing’ written as ‘sing’ (ibid.). The next example shows misspelling in the medial position, such in case when the word ‘disturb’ is spelled as ‘disturb,’ and the last example shows the misspelling in the final position: the word ‘conclude’ is spelled as ‘conclud’ (ibid.).

Orthographic errors may also be committed by mishearing words. For example, in cases when people watch television and mishearing words, such as ‘Wednesday’ is spelled as ‘Wensday’ (ThambiJose, 2014:2442). A common reason why students have problems with orthographic errors is the prevailing use of computers, when the computer automatically corrects most of the errors. Basically, to reduce the number of orthographic errors, it is

important to learn the basics of spelling rules and investigate borrowed words. Derivational morphology will also help to reduce the number of errors (ibid.).

According to Lexbook dictionary, orthography includes such rules as ‘spelling, hyphenation, capitalization, word breaks, emphasis, and punctuation’ (Online 2).

1.3.1 Spelling errors

Wachowicz and Scott (1999) have found that the spelling rules in the English language are difficult and irregular because of the phonological rules (as quoted in Al-Sauldi, 2013:100). This is the most common reason of spelling errors: phonetic errors. This type of errors occurs when the spelling of the word is incorrect, though the pronunciation is right (Vicedo, 2004:373). The words in the English language sound different in comparison with its spelling (Al-Saudi, 2013:100). Another reason of spelling errors in the English language is that students follow the rules of other languages to spell in English (ibid.). Some students have problems with /p/ and /b/ or /f/ and /v/ because, for some language learners, they sound very similar (ibid.).

According to Salam and Stevens (2007) and Watson and Wood (2002), spelling errors can be divided in two types. The first type is related with incorrect typing of words that can be caused by wrong finger movements when computer-processing. The other type deals with the knowledge of right word spelling (Salam and Stevens, 2007:386-187; Watson and Wood, 2002:10). The spelling errors that are committed due to wrong typing, for example, omission or wrong order of letters, or too many letters are defined as typographical errors (ibid.). Spencer and Edwards (1997) consider that typographical errors do not involve the knowledge of the learner but the accuracy of writing (Spencer and Edwards, 1997:5). Spencer and Edwards present the most common typographical errors: transposition of letters, absence of a necessary capital letter, no space between words or too much space, omission or addition of characters, numbers in words (because of wrong finger positioning on the keyboard), and misspelling due to wrong letter usage (ibid.).

Peters (1985) considers that the reason why students make spelling errors is because of weak memory that is caused by the use of automatic spell checkers during their essays and other tasks (as quoted in Al-Saudi, 2013:100). They start using spell checkers and are not interested in developing spelling skills (ibid.). Watson and Wood state that every written work should be checked in order to avoid these spelling errors (Watson and Wood, 2002:10).

1.3.2 Hyphenation errors

For many students, hyphenation may cause problems. Paul Brians presents the basic rules to avoid errors in hyphenation (Online 3). According to Brians, hyphenation is never used in cases when there is a combination of an adverb that ends with *-ly* and an adjective; for example: *‘His necktie reflected his generally grotesque taste’* (ibid.). Yet, other combinations of adverbs and adjectives are hyphenated, for example, *‘His long-suffering wife finally snapped and fed it through the office shredder’* (ibid.). Brians explains the point of this case: *‘Long modifies suffering, not wife;’* he also adds that words are not hyphenated when they *‘modify the same noun’* (ibid.). When both adjectives modify the same word, there is no hyphen, e.g., *‘light green suitcase’* (ibid.).

When there is a combination of adjectives and nouns that have the ending *-ed*, the hyphen is used (ibid.). For example, *‘Frank was a hot-headed cop’* (ibid.). Adjective phrases that involve a unit of measurement usually have hyphens, e.g., *‘ten-year-old car’* (ibid.).

Brians claims that *‘fractions are almost always hyphenated when they are adjectives,’* e.g., *‘He is one-quarter Irish and three-quarters Nigerian;’* however, fractions that are used as nouns are written without the hyphen, e.g., *‘He ate one quarter of the turkey’* (ibid.).

According to Brian, if there is a *‘noun and a present participle’* with the ending *-ing*, that phrase should be written with the hyphen, e.g., *‘thrill-seeking teenagers’* (ibid.).

1.3.3 Capitalization errors

According to Ritter (2003), capitalization in the English language is used to *‘punctuate sentences and to distinguish proper nouns from other words’* (Ritter, 2003:71).

Melissa Donovan says that correct capitalization plays an important role in good grammar (Online 4). She claims that nowadays people capitalize words carelessly (ibid.). There are a lot of cases when there are clearly formulated grammar rules to make the capitalization process easier; however, in some cases, the rules are unclear, such as in cases like title capitalization (ibid.). Cases where words should be capitalized are:

- Titles
- Acronyms
- First word of sentence
- Proper nouns
- Words *Web* and *Internet*
- Job titles (Online 4)

According to Donovan, all the words in titles should be capitalized, except small words, e.g., *a, an, the* (ibid.). Proper nouns should also be capitalized; for example, specific people, places or things (ibid.). As Donovan points out, there are specific words that should also be capitalized. Job titles should only be capitalized when they are titles related to a specific person (ibid.).

1.3.4 Punctuation errors

According to Davids (2005), punctuation is ‘not some optional extra in a piece of writing to be added if one chooses, but is an integral part of what is written’ (Davidson, 2005:viii). He explains that writing is different from speech (ibid.:5). When people speak they can control the speech by using different intonations, pausing or even using specific rhythms; however, in writing, the only thing of expressing meaning is punctuation (ibid.).

What is more, according to Pat Byrd, there are some punctuation errors that are typical for ESL/EFL learners (Online 5). There are difficulties in understanding such cases as *comma splices* and *fragments* (ibid.). Pat Byrd explains that a comma splice is ‘the mis-use of a comma to make a compound sentence’ (ibid.) He adds that at least two simple sentences make a compound sentence and those sentences ‘are combined to make a new sentence’ (ibid.).

There are several ways to create compound sentences (ibid.). The first option is combining two simple sentences ‘by using a comma and one of the coordinating conjunctions’ (ibid.). Byrd gives an example: ‘*He loves sociology, and he plans to major in that disciplinary area*’ (ibid.).

The second option is combining two sentences ‘by using a semicolon’ (ibid.). For example, ‘*He loves sociology; he plans to major in that disciplinary area*’ (ibid.). Byrd says that a comma splice ‘makes the connection but leaves out the coordinating conjunction,’ e.g., ‘*He loves sociology, he plans to major in that disciplinary area*’ (ibid.). Basically, it means that a comma splice is the combination of two compound sentences without any ‘coordinating conjunction’ (ibid.).

Furthermore, Byrd states that ‘a fragment is a piece of a sentence used as if it were a complete sentence,’ e.g., ‘*He loves sociology. Because he is fascinated with understanding how people behave in social groups*’ (ibid.). The example presented above shows that the second part of the current example is presented as the full sentence; however, it is only an adverbial clause (ibid.).

Concluding the exploration of orthographical error types, it can be said that they play a significant role in clarifying the meaning of the context and help to avoid misunderstandings between the writer and the reader.

1.4 Interpretation of error analysis

There are several causes for committing errors in the process of second language acquisition, and one of them is interlanguage. This term was proposed by Selinker (1972) (discussed in Khansir, 2012:1030). Before defining the term *interlanguage*, it is relevant to look at how the learners' L1 is formed. According to Selinker (1974),

We assume that there is such a psychological structure and that it is latent in the brain, activated when one attempts to learn a second language. The concept of latent language structure, which is an already formulated arrangement in the brain; is the biological counterpart to universal grammar; is transformed by the infant into the realized structure of a particular grammar in accordance with certain maturation stages. (Selinker, 1974:33)

Selinker states that interlanguage is 'a separate linguistic system based on the observable output which results from a learner's attempted production of a target language norm' (Selinker, 1974:35), and defines it as improving the skills in the second-language learning process. Second language learners who continue developing the knowledge of L2 have introduced this type of language (Online 4). He suggests that interlanguage is 'the intermediate stages between the native and target language observable in learners' language' (as quoted in Richards, 1975:29). Corder (1973) names interlanguage as 'idiosyncratic dialect;' he says that it can also be called a transitional dialect due to its instability (Corder, 1973:162). Basically, this dialect is a type of language that is spoken by a L2 learner (ibid.).

According to Ellis and Barkhuizen (2005), there are several interlanguage characteristics:

- consists primarily of implicit linguistic knowledge;
- has a grammar system just like the one of an L1 or L2;
- can be easily penetrated by new linguistic forms;
- is transitional and the learner restructures and develops it over time;
- is variable and these variations can be random in part;
- is a product of general learning strategies;
- may fossilize. (Ellis, Barkhuizen, 2005:54-55)

Moreover, Tarone (1976) has also pointed out the characteristics of interlanguage production (discussed in Khansir, 2012:1030). First, L2 speakers usually do not reach the

language performance level that is expected of L1 speakers in the same language. Second, interlanguage is not a direct translation of the speaker's mother tongue. Third, interlanguage does not occur randomly; and last, 'interlanguages are spoken either by adults or by children when second language acquisition is not simultaneous with that of the first language' (ibid.).

Selinker (1972) maintains that second language learning should be based on actual communication in L2 (as discussed in ibid.). He considers that there is 'a psychological structure latent in the brain, activated when one attempts to learn a second language' (as quoted in Richards, 1975:29). There are certain factors that explain the process of language learning and bring it to success. Richards (1975) claims that those factors could be motivation or the time available (as discussed in ibid.).

According to Selinker, interlanguage is a different system of language if compared with mother tongue (ibid.). He utilized the term 'fossilization' in order to describe the commonplace situation when L2 learners 'stop developing their interlanguage grammar in the direction of the target language' (Khansir, 2012:1030). Selinker refers to the five following concepts of fossilization, such as:

1. Language Transfer
2. Transfer of Training
3. Strategies of Second Language Learning
4. Strategies of Second Language Communication
5. Overgeneralization of the Target Language Linguistic Materials. (as quoted in Khansir, 2012:1030)

According to Selinker, the explanation of *language transfer* is that there are cases when interlanguage 'may result from the first language' (ibid.). He says that *transfer of training* is a process when there are 'some elements of the interlanguage may result from specific features of the training process used to teach the second language' (ibid.).

Selinker states that *strategies of second language learning* are 'some elements of the interlanguage may result from a specific approach to the material to be learned' (ibid.). He describes *strategies of second language communication* stating that 'some elements of the interlanguage may result from specific ways people learn to communicate with native speakers of the target language' (ibid.). *Overgeneralization of the target language linguistic materials* is explained with the idea that 'some elements of the interlanguage may be the product of overgeneralization of the rules and semantic features of the target language' (ibid.).

Richards (1975) mentions the importance of other processes as spelling pronunciation, cognate pronunciation, holophrase learning, hypercorrection and others (Richards, 1975:41). Nemser (1971) uses a different terminology. Instead of interlanguage, he uses the term *approximative system*. As the learners' L2 systems are continuously reforming because of the

elements that are being added to the learning process, this term may be considered more applicable than *interlanguage* due to ‘developmental nature of language learning’ (ibid.:30). Nemser mentions that those developmental systems are obvious in the errors of learners (ibid.).

Although the above presented information sets a stable foundation for L2 acquisition research, language learning psychology is a highly complex part of the research. Selinker (1972) considers that the most difficult part in second language learning psychology development is the scholars’ hardship of recognizing and clearly defining the methods and expected outcomes of finding out the exact ways how L2 learners acquire the language (ibid.:32). There are set of behavioural events ‘which has elicited considerable interest is the regular reappearance in second language performance of linguistic phenomena’ which were thought to be eradicated in the performance of the learner’ (ibid.).

The following subchapter focuses on the corpus-based approach of error analysis, as the present research is concerned with this method of analysis.

1.5 Corpus-based approach error analysis

According to Biber, Conrad and Reppen (1998), *structure* and *use* are the two main areas of language studies. Generally, linguistic analysis has highlighted the studies of structure by ‘identifying structural units and classes of language,’ for example, morphemes, words or phrases, and ‘describing how small units can be combined to form larger grammatical units,’ for example, how words are combined into phrases (Biber, Conrad, Reppen, 1998:1).

However, studying language from the perspective of real, everyday language highlights the language use outside of isolated grammar book texts. In corpus-based language analysis approach, the focus is not on the theoretical possibilities of the language but rather on the study of ‘the actual language used in naturally occurring texts’ (ibid.). Basically, it is a study that investigates the way language users operate with their language of use.

When attempting to analyse language use, different questions may occur, for example, ‘do spoken varieties versus written varieties have different preferences for one of the forms over others?’ or ‘are the forms usually used with different verbs?’ (ibid.).

The important part of language use analysis is ‘comparing the language of different texts or groups of texts’ (ibid.:3). Everyday people participate in different situations, such as reading a magazine or talking to a friend, and this leads to the varieties of language use, called *registers* (ibid.:3).

When investigating language use, researchers try not to judge grammaticality but rather ‘uncover typical patterns’. Analyses of language use have ‘two central research goals’ (ibid.). The first goal is ‘assessing the extent to which a pattern is found,’ and the second goal is ‘analysing the contextual factors that influence variability’ (ibid.). There are difficulties in analysing large amounts of language, presented by speakers and writers. To facilitate such difficulties, a corpus-based approach ‘provides a means of handling large amount of language and keeping track of many contextual factors at the same time’ (ibid.). Basically, it can be said that corpus-based approach has helped researchers by facilitating language use investigation (ibid.).

Biber, Conrad and Reppen identify the characteristics of corpus-based analysis. They underline these characteristics as a new perspective on language use: ‘studying the use of language characteristics by considering the relevant ‘association patterns’’ (1998:4).

The characteristics of corpus-based analysis are:

- it is empirical, analyzing the actual patterns of use in natural texts;
- it utilizes a large and principled collection of natural texts, known as a corpus; as a basis for analysis;
- it makes extensive use of computers for analysis, using both automatic and interactive techniques;
- it depends on both quantitative and qualitative analytical techniques. (Biber, Conrad, Reppen, 1998:4)

Using computers brings advantages to the corpus-based approach because computers identify and analyse complex language patterns and deals with the analysis of a large database. Computers are also used for record-keeping while analysts are making linguistic judgements (ibid.).

It is important to admit that corpus-based analysis must include ‘qualitative, functional interpretations of quantative patterns’ (ibid.:5).

Furthermore, Biber, Conrad and Reppen claim that a corpus-based approach gives an opportunity for researchers to recognize the ‘association patterns’ and they explain these patterns as ‘the systemic ways in which linguistic features are used in association with other linguistic and non-linguistic features’ (ibid.).

In terms of association patterns, there are two research questions (ibid.). The first one is ‘to focus on linguistic features;’ it can be focused either on a ‘lexical item or grammatical construction’ (ibid.). And the second aspect is ‘to focus on the characteristics of texts or varieties’ (ibid.).

Biber, Conrad and Reppen maintain that, with corpus-based studies, many kinds of association patterns can be investigate. They point out that those patterns ‘represent

continuous relationships.’ It means that the patterns have something in common and they ‘occur to differing extents’ (ibid.:8).

Quantitative techniques are also relevant for corpus-based studies. For example, if there was a need to compare the patterns of language use for two similar words, the following aspects would be researched: the frequency of the particular words in the corpus, how many different collocations occur and how common they are (ibid.).

Corpus-based studies are concentrated on explaining large, quantitative patterns, and that takes a significant amount of effort (ibid.:9).

Biber, Conrad and Reppen continue by noting that ‘corpus-based analysis should be seen as complementary approach to more traditional approaches, rather than as the single correct approach’ (ibid.).

According to Biber, Conrad and Reppen, language use can be studied ‘through detailed analyses of specific linguistic features in particular texts, complementing the findings from analysis of large corpora’ (ibid.:10).

The corpus-based approach can be applicable to various topics in linguistics: they can concentrate on individual words, on different grammatical features, on styles of the author or register patterns (ibid.:11).

Lastly, Biber, Conrad and Reppen say that corpus-based techniques can be applied to lexicography and grammar. They give an example: earlier, the work of dictionary makers was limited and their task was to identify the meaning of a word, whereas now they have an option to provide information on the meaning used more frequently and the contexts in what the words are used (ibid.).

Besides the positive characteristics of the corpus-based research method, it is relevant to admit that several researchers have criticized this approach. Widdowson (2000) claims that ‘corpus linguistics only offers a partial account of real language,’ and it is explained by it not addressing ‘the lack of correspondence between corpus findings and native speaker’s intuitions’ (as quoted in Baker, 2006:7).

Borsley and Ingham (2002) criticize the language conclusion making process within the corpus-based approach because the nature of a corpus is not sufficient to reflect the true language use intentions. They argue that ‘language is endowed with meaning by native speakers and therefore cannot be derived from a corpus’ (ibid.). Baldry (2000) also criticizes this approach because it separates texts from their contexts (as discussed in ibid.).

As regards error analysis, Biber (1998) claims that it is considered to be useful to apply corpus-based techniques in second-language investigation. Errors are an area that always causes a discussion among second-language teachers. Errors made by L2 students are looked

at from different perspectives. Biber says that those perspectives include ‘the gravity of different kinds of errors, the nature of errors as interlanguage and the stages by which errors are replaced by target forms’ (Biber, 1998:197). In reality, teachers may feel annoyed because of the local errors that do not change the meaning of the sentence; for that reason, students may see it in the progress evaluation process.

1.5.1 Defining corpus

According to Flowerdew (2012), Corpus Linguistics is defined as ‘a collection of authentic language, either written or spoken, which has been compiled for a particular purpose’ (Flowerdew, 2012:3). She states that the purposes are of either purely linguistic or socio-pragmatic nature (ibid.).

Curzan and Palmer (2006) claim that corpus (from Latin *body*) is ‘a large and principled collection of texts’ that is characterized by several features (as quoted in Facchinetti, Rissanen, 2006:18). McEnery and Wilson (1996) present the features that characterize a corpus: it should have ‘sampling and representativeness,’ be in ‘machine readable form,’ have ‘finite size,’ and a ‘standard reference for the language variety it represents’ (as quoted in Facchinetti, Rissanen, 2006:18).

McCarthy (2001) claims that investigation of corpora that are ‘social artefacts’ can reveal the socio-pragmatic behaviour of specific discourse communities (as discussed in Facchinetti, Rissanen, 2006:18).

These days, it is agreed that corpus data is available electronically in machine-readable form (ibid).

Flowerdew (2012) presents the criteria for defining a corpus. The first criteria is that a corpus is composed of ‘authentic, naturally occurring data’ (Flowerdew, 2012:3). As the second criteria – that a corpus is collected in accordance with precisely designed criteria. Next, that a ‘corpus is representative of a particular language or genre’ (ibid.). And, lastly, that a corpus is designed for a specific purpose: linguistic or socio-pragmatic (ibid.).

Sinclair (1991) states that a corpus is ‘a collection of naturally occurring language texts, chosen to characterize a state of variety of a language’ (as quoted in Flowerdew, 2012:3). He gives an explanation that corpora are generally many million words large, explaining that the reason of such amount of words is the creativity of a natural language that leads to such enormous variety of expression (ibid).

According to Taubert and Chermakova (2004), corpus linguistics has a different look at language. They say that language is a ‘human faculty which children acquire naturally

without being given instructions; it is set of rules we have learned from forming plural nouns, to using words in the appropriate order, to following conventions of letters or essays or reports, and it is a long list of words we have learned' they add that it is also 'the sum of all texts in that language' (Taubert and Chermakova, 2004:36).

Taubert and Chermakova state that 'corpus linguistics sees language as a social phenomenon,' saying that 'language is social phenomenon' (ibid.). Taubert and Chermakova say that 'corpus linguistics is concerned with the message itself' and it can explain the meaning (ibid.:39).

According to McEnery and Wilson (1996), corpus linguistics is 'the study of language based on examples of real life language use' (as quoted in Baker, 2006:1). Corpus linguistics applies quantitative methodology, for example choosing a linguistic phenomenon and finding out the frequency information (ibid:2).

Generally, corpora are large, consist of large number of words, represent samples of naturally occurring language, 'so they can therefore be used as a Standard reference with which claims about language can be measured' (ibid.). The benefits of electronic corpora is that they can deal with complex calculations in large amounts of text, detecting linguistic patterns and information about frequency that would normally take significantly more time doing everything by hand (ibid.).

Frequently, corpora are annotated with complementary linguistic information, usually 'the most common being part of speech information (for example, whether something is a noun or a verb), which allows large-scale grammatical analyses to be carried out' (ibid.). Other information that can be encoded is attributes such as 'sex, age, socio-economic group and region' of spoken corpora, that contains transcripts of dialogue. It allow for language comparisons of different speakers (ibid.).

There are number of areas of linguistic enquiry where corpus linguistics has been used; for example, in the creation of dictionaries, 'an aid to interpretation of literary texts, forensic linguistics, language description, language variation studies and language teaching materials' (ibid.:3).

According to Gilquin and Gries (2009), 'corpora are a supreme tool for linguistic data analysis' (as quoted in Cigankova and Vinchela,2012:246). They state that using the data that is 'coming from natural contexts' gives an opportunity to investigate the cases that previously were unexplored (ibid.). Cigankova and Vinchela add that 'the application of corpora make such studies more objective and raises external validity of research' (ibid.:247).

According to Sinclair (2004) a corpus is a considerable thing not only because 'it is a collection of language text' but also because of its properties if the corpus is well-designed

and carefully-constructed (Sinclair, 2004). Principles that are related to corpus and text are difficult to define; however, there are things that need to be faced in corpus design such as ‘sample’ and ‘balance,’ and an unavoidable one - ‘the notion of representativeness’ (ibid.).

1.5.2 Types of corpora

To compile a corpus, it is important to understand the types of corpora that presented by researchers Sinclair (2004), Baker (2006), Meyer (2002) and others.

Even though the term *corpus* means ‘a body of electronically encoded text,’ it does not mean that those texts are selected at random (Baker, 2006:26). Researchers have presented several types of corpora depending on the research goals (ibid.).

The first type is *specialized corpus*. According to Baker, this corpus type would be used ‘in order to study aspects of a particular variety of genre of language’ (Baker, 2006:26). Basically, that means that if the research question is to analyse the language of magazines, texts that correspond to this specialized criteria will be collected. There are possibilities to specialize in a specific topic, for example, Johnson et al. (2003) have ‘built the corpus of British newspaper texts that contained references to the concept of political correctness’ (as quoted in Baker, 2006:26). His criteria of inclusion was to select the articles that contain phrases like ‘politically correct, PC or political incorrectness’ (ibid.).

According to Leech (2002), Sinclair (1982) has developed the relevant idea by inventing the term *monitor corpus* (Leech, 2002:2). This idea presents the fact that it is not enough to build a corpus and finish it, but it is sufficient to monitor the changes in language (ibid.). Leech exemplifies: ‘what new words are coming to use, what old words disappear, how the grammatical structures evolve, morphology evolves, and so on’ (ibid.). Basically, it means that new information should be added to the corpus to follow how the language is developing (ibid.).

Recently, such corpora as *parallel corpora* have been developed (ibid.). This type of corpora deals with translation (ibid.). Leech explains that in this corpus, there are two sets of texts in different languages (A and B) (ibid.). Having this data makes it possible to build the corpora of different languages, allowing for the opportunity to compare between more than two languages (ibid.).

Another recent type of corpora is *comparable corpora*, which is a little different from parallel corpora. It means that at least two corpora are built, applying the same principles of data selection (ibid.). In that case, it is possible to compare different varieties via corpora (ibid.).

The next type, *diachronic corpus*, is built ‘in order to be representative of a language variety over a particular period of time, making it is possible for researchers to track linguistic changes within it’ (ibid.:29). Baker (2006) gives an example of a diachronic corpus – the Helsinki Corpus of English Texts (2006). The data for this corpus was collected from diverse written texts from the earlier periods of English Language (Meyer, 2002:xiii). Meyer (2002) says that this corpus has been developed to investigate the evolution of English by historical linguists (Meyer, 2002:xiii).

The following type of corpus is *reference corpus*. According to Baker (2006), this type of corpus is ‘what purists would generally refer to when they use the term corpus’ (Baker, 2006:30). The example of such corpus is the British National Corpus that contains one hundred million words from various ranges of texts (ibid.). This corpus consists of texts of ‘extracts from regional and national newspapers, specialist periodicals and journals, academic books and popular fiction, published and unpublished letters and school and university essays’ (ibid.). Reference corpora can help in testing different hypotheses; for example, by looking at the frequency of certain words (Baker, 2006:43).

As the aim of the current BA thesis is to build the corpus of The University of Latvia students’ essays, the corpus type for the empirical part will be reference corpus.

According to Geoffrey Leech, this corpus is very important because it is designed in order to ‘provide comprehensive information about the language’ (Leech, 2002:1). Moreover, this type of corpus ‘has to be accepted as a de facto standard for the language and users have to feel that it somehow represents the language’ (ibid.:3). It is important to make the reference corpus as representative as possible, ‘so it will be accepted by the community of speakers of the language, and particularly those studying the language, as some kind of standard reference point’ (ibid.).

To achieve the representativeness in corpora, three factors that have to be achieved exist (ibid.:5). The first factor is diversity; Leech explains: ‘a full sampling of the varieties of the language, as wide a range as possible of the varieties of the language’ (ibid.). The next factor is balance; Leech defines it as ‘the subsamples or the subcorpora of different language varieties must in some sense be proportionate to their importance in the language,’ and the last factor is size, which is not as important as diversity and balance, but Leech claims: ‘the more data always better’ (ibid.).

Chapter summary

The first chapter consists of two parts, where the first part deals with the error analysis approach, presents the theoretical background of the error analysis, shows what types of errors exist in L2 learning process, investigates particularly orthographic errors and reveals their types. This part of the chapter also deals with the analysis of error causes. The second part of the chapter deals with the corpus-based approach error analysis; it gives an insight into the characteristics of corpus-based approach, explains what corpus is and presents the types of it.

2 RESEARCH METHODOLOGY

This chapter is devoted to the description of the methodology and the procedure of the current BA paper. Firstly, the corpus, which created by the author of the BA thesis, is described and characterized, and the principles of how it was created are presented. Secondly, this chapter presents the steps on how the research was done. The research contains a method of spotting the frequency of orthographic errors (quantitative analysis), as well as qualitative research is carried out in order to provide a more detailed analysis of the essays written by the students. The following subchapters will present a more detailed description of the analysis.

2.1 Research method

The current BA thesis contains two research methods: a corpus-based approach and an error analysis approach. The corpus-based approach includes a quantitative method that investigates the frequency of orthographic errors in the essays written in the classroom. The corpus of the analysis includes 59 essays in electronic form compiled by the author of the research.

The error analysis approach includes a qualitative method of analysing the orthographic errors made by students in their essays written in the classroom. In order to carry out an error analysis, the following steps, presented by Gass and Selinker (2008), were put into practice. They were as follows:

1. Collection of a sample of learner language
2. Identification of errors
3. Classification of errors
4. Frequency of errors
4. Explanation of errors
5. Evaluation of errors.

(Gass and Selinker, 2008:103)

59 essays were collected and a corpus was compiled in order to follow the first step. Secondly, the errors were identified to understand their nature. In accordance with the following step, the orthographic errors were classified according to their type (spelling, capitalization, hyphenation, punctuation and syllabication errors). As the next step, the errors were counted to understand the quantity of each orthographic error. And, finally, the errors were explained and evaluated by whether they were local or global errors (Kiparsky, 1974).

2.2 Research object

The current study involves corpus linguistics and analyses the most common orthographic errors in the students' essays and investigates their frequency. Since the empirical part of the current BA thesis investigates the certain error aspects, the corpus that was compiled for the study is further considered as 'reference corpus' (Baker, 2006:26).

The research object of the current research is the reference corpus that was compiled particularly for this study. It consists of first-year students' essays from the University of Latvia. The length of each essay was approximately 600 words, however the total number of words was 28 290. The essays were written by the students who study in the Modern Language and Business Study programme; however, the essays were selected at random and they did not provide any information about the students. Students' have written their essays in the classroom in Business English, and the topic of the essays was related to business. The essays were word-processed and the students were able to use spell checkers.

The corpus consists of 59 essays that was collected and analysed in order to investigate not only the orthographical errors but also to look at the other errors that were committed by the students. All the essays were written in the English language and were presented in electronically written form. The length of each essay was approximately 600 words. The authors of the essays were first-year students learning English as the second language who had studied Academic Writing and their native language was various, i.e. Latvian or Russian.

2.3 Research procedure

To reach the goal of the BA paper - to investigate the frequency of orthographic errors and to identify their most common types in students' essays written in Business English -, certain procedures were carried out.

In order to reach the goal of the research, firstly, the author of the BA thesis has compiled a corpus of 59 essays written by the students of Business English in the University of Latvia. As the next step, correction of all orthographical errors found in the essays was done. To investigate the frequency of the orthographical errors, they were counted. Moreover, in order to find out the most common type of them, the errors were classified by their type (spelling, capitalization, punctuation, syllabication, hyphenation errors). In the essays, examples of sentences that contain orthographic errors were found. Then, with the help of *Microsoft Excel*, a frequency analysis of each orthographic error type was done.

After identifying and analysing the orthographical mistakes, the results and research findings with real examples, and comments on the possible causes of such mistakes were presented. Finally, relevant conclusions were drawn.

Chapter summary

The second chapter comprises the methodology of the current research; it describes two approaches: corpus-based and error analysis approach. It describes the analysis and the methods that were used in the process of the research, and explains how the methods were applied in order to analyse the frequency and types of the errors. Also, it describes the corpus that has been compiled by the author of the current BA thesis, characterizing the written documents that were used for the corpus. Finally, it describes the procedure of the orthographical error analysis and explains how the errors found in the essays were classified. The following chapter will present the results and findings of the present research.

3 RESULTS AND DISCUSSION

This chapter describes the empirical part of the current research. It analyses orthographical errors in the corpus of 59 students' essays.

Firstly, the author of the current BA paper presents the overall results of the research. Secondly, the most common errors in students' essays are demonstrated. Thirdly, the author discusses the findings of the empirical part.

3.1 Research results

The research results present the orthographical error analysis of 59 first-year Modern Language and Business Studies students' Business English essays written in the classroom. The students' native language was mainly Latvian and Russian and they studied English as a foreign language. The analysis was done in order to identify the orthographical errors' frequency and their most common types in students' essays written in the classroom in Business English. In order to identify these errors, the author of the current research has compiled a corpus of 59 essays. Afterwards, these errors were analysed.

With the help of *Microsoft Excel*, orthographic errors were classified according to their type, whether it was a spelling error, a hyphenation error, a capitalization error, syllabication error, an error of emphasis or a punctuation error. The summary of the orthographic errors is presented in Table 3.1.

Table 3.1 Number of orthographical errors in students' essays

Type of orthographic error	Number of errors
Spelling errors	52
Hyphenation errors	31
Capitalization errors	10
Word breaks errors	0
Errors of emphasis	8
Punctuation errors	63
Total number of orthographic errors:	164

Table 3.1 presents the number of each orthographic error category. According to the table above, 164 orthographic errors were recognized in 59 students' essays.

From the 164 orthographic errors, 63 were identified as punctuation errors. Cases of wrong used punctuation or missed commas were noticed. Sentences where supplementary nonessential information was introduced to a clause were not surrounded by commas were recognized, e.g., 'The cause of *consumerism is the environment which has been developed* by the human society'. Cases when a comma was not used in sentences that start with a dependent clause were also identified, e.g., 'After Estonia joined the euro zone Foreign investment increased up to 10,4% within the year'. Also, students had problems with the comma splice. There were cases where the students had written more than one independent sentence without a coordination conjunction and without the use of a comma, e.g., 'Students are not able to work during their studies they do not have time due to the amount of the homework' and 'After graduating the school, students are expecting to find a good job'.

The author of the present BA paper has identified 52 spelling errors. These errors were different, for example, the omission of double letters, mistyping and switching letters. Table 3.2 shows some examples of the errors that were found in the students' essays.

Table 3.2 Examples of spelling errors in students' essays

Error	Correct Spelling	Comments
illegal	illegal	The root of the word is <i>legal</i> and the prefix is <i>-il</i> .
can not	cannot	<i>Cannot</i> is an admissible form; however, <i>can not</i> is used when there is another construction with <i>not</i> , for example, <i>not only</i> (Online 12), e.g., The company <i>can not only</i> provide their employees with salary but also ensure benefits.
motive	motivate	In the context of the sentence, the word <i>motivate</i> is more appropriate, e.g., The employees feel <i>motivated</i> if they receive benefits.
avaliable	available	Common misspelling of the word <i>available</i> .
comperes	compares	Misspelling due to pronunciation of <i>-a</i> that is /ə/.
standarts	standards	Misspelling of the word because of the influence of the first language (in Latvian <i>standarts</i>).
team work	teamwork	The word <i>teamwork</i> is used in order to express the cooperation of people to get results, e.g., Many employees are involved in the <i>teamwork</i> to reach the goal

		of the company.
work place	workplace	<i>Workplace</i> is written as one word as it is meant to be <i>office</i> .
carrier	career	In some contexts, <i>career</i> is appropriated because carrier, according to Oxford Dictionaries, is ‘A person or thing that carries, holds, or conveys something’ (Online 6); however, <i>career</i> is ‘An occupation undertaken for a significant period of a person’s life and with opportunities for progress’ (Online 7), e.g., His <i>career</i> depends on his education.
then	than	In comparisons, conjunction <i>than</i> should be used instead of adverb <i>then</i> , e.g., Some companies are doing better <i>than</i> others.
than	then	If there is no comparison in the sentence, and the adverb <i>then</i> should be used because of the meaning of the adverb in this sentence is to show the following action in the future, e.g., Some people start as administrators then become managers.
per cents	per cent	<i>Per cent</i> is always used in singular.
programm	program/ programme	There are two options to say <i>programme</i> or <i>program</i> ; however, <i>program</i> is more used in IT sphere (Online 11).
sometime	sometimes	In the current sentence, the word should be used with the meaning of ‘not always or not usually’ (Online 10), e.g., <i>Sometimes</i> employees are not satisfied with their jobs.
worls	world	Misprint on the keyboard (as–s and –d are located next to each other).
finaly	finally	Common misspelling; according to Oxford Dictionaries, <i>finally</i> should be spelled with double –l (Online 9).
diner	dinner	In the context, <i>dinner</i> is used as an event, e.g., Employees are having <i>dinner</i> together.
tuitions fees	tuition fees	The first word should be in singular and the second in plural.
life style	lifestyle	Is written as one word.
raise	rise	Misspelling because of the pronunciation of –i that is /aɪ/ in the word <i>rise</i> .
wether	whether	According to the Oxford Dictionary of Etymology, <i>wether</i> is ‘male sheep or ram’ (as cited in Online 8); however, <i>whether</i> is a conjunction that would have been appropriate in the sentence, e.g., Students have problems to decide whether study or not.
never the less	nevertheless	The adverb is spelled as one word.

Chesteron	Chesterton	Misspelling of the surname of a writer and philosopher.
consmerism	consumerism	Negligence error.
futhemore	furthermore	The root of the word is <i>further</i> .

The table above shows the errors that were found in the students' essays. It is possible to foresee the reason why those errors occur in the essays. The first reason is negligence: in some cases, one character has been missed, for example, instead of *consumerism*, it is written as *consmerism*. This kind of errors may occur due to fast typing or a wrong position of the fingers on the keyboard. The next possible reason is a phonetic error. In some cases, it is difficult to spell the word correctly because of the pronunciation of it, for example, the word *rise* has been written as *raise* by the student, only because it is pronounced in this way when speaking. Another possible reason is the lack of knowledge, as in the example of *never the less*. The student has divided one word into three; it might have happened because the student simply did not know the right spelling of it. The following reason is the influence of the mother tongue, for example, the word *program* (American English) or *programme* (British English). Several students have written it with the double *-m* and without the *-e* at the end of the word (*programm*), so that the word does not match neither of the spelling variants. It has probably occurred because the native language of the students were Latvian and in this language the word is written as *programma*.

One more possible reason is the lack of understanding; for example, when *than* is used instead of *then* and vice versa. It might have occurred because the students simply do not understand that these two words are used in different situations.

During the analysis, 10 cases of different capitalization errors were identified. The reason of these errors may also be different, for example, inattentiveness or the lack of knowledge. Table 3.3 shows some of the examples of errors in capitalization.

Table 3.3 Examples of capitalization errors in students' essays

Error	Correct Capitalization	Comments
'the traveller sees what he sees'	'The traveller sees what he sees'	The first sentence of full quotation was not capitalized.
tower of London	Tower of London	Proper names should be capitalized, for example, <i>Buckingham Palace</i> , <i>Trafalgar Square</i> .
'if in the shop was only one or two	If in the shop was only one or two lawnmowers, consumer	Capitalization of the first letter in a new sentence.

	had to buy one of these two...	
‘After Estonia joined the euro zone Foreign investment increased up to 10,4% within the year’	After Estonia joined the euro zone foreign investment increased up to 10,4% within the year’	The word <i>foreign</i> is capitalized in the middle of the sentence; however, this word is not considered as a proper noun.
For instance, in December of 2011 Latvia’s export was only 490 Million Lats.	For instance, in December of 2011 Latvia’s export was only 490 million lats.	Currencies in the English language are not considered as proper nouns. In the given example, <i>Million Lats</i> should not be capitalized.

31 hyphenation errors were identified. There were cases when hyphens were missed or used inappropriately. As the analysis shows, hyphenation seemed to be the most confusing part for the students, as well as punctuation. Some examples of hyphenation errors in students’ essays are available in Table 3.4.

Table 3.4 Examples of hyphenation errors in students’ essays

Error	Correct Hyphenation	Comments
well known locations	Well-known locations	In the sentence, <i>well known</i> is used as an adjective; however, the hyphen is missing.
‘The export and import situation is looking to improve as it has, year-by-year’	The export and import situation is looking to improve as it has, year by year.	<i>Year by year</i> is used as a period of time, not as an adjective; for that reason, it should not be hyphenated.
‘To conclude, it is very important to make white collar workers motivated in Latvia’	To conclude, it is very important to make white-collar workers motivated in Latvia	<i>White-collar</i> should be hyphenated because it describes the specific kind of worker and not a white collar.
‘And that would be considered the breaking-point for most’	And that would be considered the breaking point for most.	The hyphen is not needed in the word <i>breaking point</i> as it is not used as an adjective.

3.2 Discussion on findings

The findings show that the most common types of orthographical errors in students' essays are spelling and punctuation errors. The majority of the students used too long sentences that resulted in having punctuation errors. Another significant problem is the comma splice: students do not divide independent sentences in two or more parts, and that leads to errors.

Spelling errors in students' essays are also common, even though there were allowed to use spellcheckers while writing their essays.

During the analysis of the essays, many grammatical errors were identified. They show that the students had problems with the use of articles, prepositions, and even tenses.

These findings lead to the conclusions presented in the following part of the thesis.

CONCLUSIONS

Students have to write a significant amount of essays in the English language during their studies in the Modern Language and Business Studies programme. For the majority of the students, English is their second language. For that reason, it is relevant to investigate their abilities to write orthographically correct in English as a second language. The goal of the current BA thesis was to investigate the most common orthographic errors in students' essays with the help of a corpus compiled of Business English essays written by the students of the University of Latvia, as well as to determine what type of orthographical errors are most common in such texts.

Summarizing all the results of the current research, it is possible to say that there is a large number of orthographic errors in the students' essays. This finding shows that the first-year students have to improve their writing skills. Among the most frequent type of errors were orthographic errors such as spelling and punctuation errors; however, capitalization and hyphenation errors were found as well. Even though the students were able to use a spell checker during the essay writing, they had committed a notable number of errors. The reasons of these errors are different: in some cases, they were caused by the lack of knowledge, pronunciation that was different to the spelling, or the influence of the students' native language. Sometimes errors had occurred due to mistyping or simply negligence. It is relevant to observe that these essays were written in the classroom, and the reason of the errors may be connected with the lack of time during the lecture. However, there were many punctuation errors that showed that the first-year students simply did not have enough experience to apply the correct punctuation.

The majority of errors found during the analysis were local errors that do not usually change the meaning of the sentence and do not cause misunderstandings. Nevertheless, first-year students appeared to not know how to apply the orthographic rules in written tasks, even though the students had experience in writing essays from secondary and high school. It can also be added that students have to improve their grammar skills in order to write better essays. To use the second language without errors, the learners should understand the system of the language and see the differences between each language in order to result in using appropriate sentences. Therefore, it is important for teachers to help their learners distinguish between these differences, for example, by emphasizing certain rules that are different in the languages, or to help the students understand their errors in order to reduce the number of

them in the future. It is also relevant to point out that students should be more careful when writing essays and reread their assignments before submitting them.

To conclude, the hypothesis that orthographic errors are less frequent than other types of errors made in students' essays was not approved.

The research is limited because it consists of 59 written essays of the University of Latvia first-year students' who studied Business English as part of the Modern Language and Business Studies. For that reason, further research should be done in order to investigate the orthographic errors of other students' essays on different topics to give more precise information on the most common errors committed by the second language learners of the University of Latvia. For further research, it is relevant to continue analysing students' grammatical errors and find out ways to reduce their number, because, during the present analysis, a great number of them were observed.

THESES

1. The errors made by language learners in the process of language acquisition are considered to be natural and unavoidable; their number reduces while the level of language proficiency improves.
2. To investigate the language produced by language learners, a corpus-based approach has appeared to be a useful strategy.
3. Even though students wrote the essays in the classroom, using computers equipped with spell checkers, there were orthographical errors that frequently occurred in the essays.
4. The author of the present research has identified 164 orthographic errors in 59 students' essays written under supervision in the classroom.
5. From the 164 orthographic errors identified during the analysis, 63 were recognized as punctuation errors.
6. The punctuation errors included such cases as: sentences where supplementary nonessential information was introduced to a clause were not surrounded by commas; a comma was not used in sentences that begun with a dependent clause; and where several independent sentences were put together without a coordination conjunction and lacking a comma.
7. The author revealed various types of errors, for example, the omission of double letters, switched letters, homonym confusion and typographic errors.
8. The most common orthographical errors in the essays of the students of the University of Latvia were spelling and punctuation errors.
9. The punctuation errors in the essays were sometimes caused by misunderstanding of the meaning of the sentence.
10. The majority of the orthographical errors were local errors, which did not dramatically affect the meaning of the sentences in the essays.
11. The results of the empirical research provide the evidence that the hypothesis cannot be accepted: the number of orthographical and punctuation errors is still high in students' essays, though the students use computer programmes to find errors when they are writing essays.

REFERENCES

1. Al-Saudi, J. (2013) *Error Analysis and Spelling Mistakes of EFL Learners at Tafila technical University: A Case Study*. Jordan: The World Islamic Sciences and Education University. Available from:
https://www.academia.edu/4140208/Error_Analysis_and_Spelling_Mistakes_of_EFL_Learners_at_Tafila_Technical_University_A_Case_Study
2. Baker, P. (2006) *Using Corpora in Discourse Analysis*. London: Continuum
3. Baldry, A. (2000) *Multimodality and Multimediality in the Distance Learning Age*. Campobasso: Palladino
4. Berkelt, B., Smedt, K. (1988) *Triphone Analysis: A Combined method for Correction of Orthographical and Typographical Errors*. The Netherlands: University of Nijmegen. Available from: <http://www.aclweb.org/anthology/A88-1011>
5. Biber, D., Conrad, S., Reppen, R. (1998) *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press
6. Brown, D. (1994) *Principles of language learning & teaching*. Eaglewood Cliffs, New Jersey: Prentice Hall
7. Burt, M. and C. Kiparsky. 1972. *The gooficon: A repair manual for English*. Rowley, Massachusetts: Newbury House
8. Cigankova, N. Vincela, Z. (2012) *Specialised Corpora Structural and Functional Variability in Sociolinguistic Studies of Variation in English*. University of Latvia
9. Connor, U. (1996) *Constrative rhetoric: Cross-cultural aspects of second language writing*. Cambridge: Cambridge University Press
10. Corder S.P. (1971) Idiosyncratic Dialects and Error Analysis. In Richards, J.C. (ed.) (1974) *Error Analysis. Perspectives on Second Language Acquisition*. United Kingdom, London: Longman Group Ltd. (pp.158-171)
11. Corder, S.P. (1967) *The significance of learners' errors* in Richards, J.C. (ed.) (1974) *Error Analysis. Perspectives on Second Language Acquisition*. United Kingdom, London: Longman Group Ltd. (pp.19-30)
12. Corder, S.P. (1973) *Introducing Applied Linguistics*. Harmondsworth: Penguin Education
13. Curzan, A., Palmer, C. (2006) The importance of historical corpora, reliability and reading. In Facchinetti, R., Rissanen, M. (2006) *Corpus Based Studies of Diachronic English*. Germany: Peter Lang
14. Davidson G. (2005) *How to Punctuate*. England: Clays Ltd

15. Dulay, H.C. Burt, M.K. and Kreshen, S. (1982). *Language Two*. New York: Oxford University Press
16. Ellis, R.; Barkhuizen, G. (2005) *Analysing Learner Language*. United Kingdom, Oxford: Oxford University Press
17. Erdogan, V. (2005) Contribution of Error Analysis to Foreign Language Teaching. *Mersin University Journal of the Faculty of Education*, 2: 261-270.
18. Facchinetti, R., Rissanen, M. (2006) *Corpus Based Studies of Diachronic English*. Germany: Peter Lang
19. Ferris, D.R. (2005) *Treatment of error in second language writing*. Ann Arbor: The University of Michigan Press
20. Fisiak, J. (1981) *Contrastive Linguistic and the Language Teacher*. Oxford: Oxford Pergamon Press
21. Flowerdew, L. (2012) *Corpora and Language Education*. United Kingdom: Palgrave Macmillan
22. Gefen, R. (1979) The analysis of pupils' errors. *English Teachers' Journal*, 22: 16-24.
23. Gilquin, G., Gries, S. (2009) Corpora and experimental methods: a state-of-the-art review. In: *Corpus Linguistics and Linguistics Theory* (2009), 5(1): 1-26.
24. Jain, M. P. (1974) Error Analysis: Source, Cause and Significance. In Richards, J.C. (ed.) (1974) *Error Analysis. Perspectives on Second Language Acquisition*. United Kingdom, London: Longman Group Ltd. (pp.189-215)
25. James, C. (1998) *Errors in language learning and use. Exploring error analysis*. Essex: Pearson Education Limited
26. James, C. (1998) *Errors in Language Learning and Use*. United Kingdom: Longman
27. Johnson, S., Culpeper, J., and Suhr, S. (2003) From 'politically correct councillors' to 'Blairite nonsense': discourse of political correctness in three British newspapers. *Discourse and Society*, 14(1): 28-47.
28. Kasper, G., Kellerman, E. (1997) 'Introduction', in G. Kasper and E. Kellerman (eds.), *Communication Strategies: Psycholinguistic and Sociolinguistic Perspectives*. London: Addison Wesley Longman, pp. 1-13.
29. Khansir, A.A. (2012) *Error Analysis and Second Language Acquisition*. Finland: Academy Publisher. Available from:
<http://www.ojs.academypublisher.com/index.php/tpls/article/viewFile/tpls020510271032/4942>

30. Leech, G. (2002) *The importance of reference corpora*. Lancaster University.
Available from:
<http://www.corpus4u.org/forum/upload/forum/2005060301260076.pdf>
31. Mc Carthy, M. (2001) *Issues in Applied Linguistics*. Cambridge University Press
32. McEnery, T, Wilson, A. (1996) *Corpus Linguistics: An Introduction*. Edinburgh University Press
33. Meyer, C. (2002) *English Corpus Linguistics*. United Kingdom: Cambridge
34. Norish, J. (1983) *Language Learners and their errors*. London: Modern English Publications
35. Peters, M.L. (1985) *Spelling Caught or taught? A New Look*. London: Routledge
36. Richards J.C. (1975) *Error Analysis. Perspective on Second Language Acquisition*. United Kingdom: Longman Group Limited
37. Richards, J.C. (1971) A Non- Contrastive Approach to Error Analysis. *Journal of ELT*, 25: 204-219.
38. Ritter, R. M. (2003) *Oxford Style Manual*, United Kingdom: Oxford University Press
39. Selinker, J. (1974) Interlanguage. In Richards J.C. (1975) *Error Analysis. Perspective on Second Language Acquisition*. United Kingdom: Longman Group Limited
40. Selinker, L. (1972) Interlanguage. *IRAL*, 10: 209-231.
41. Selinker, L., Gass, S.M., (2008) *Second Language Acquisition*. New York: Routledge
42. Sharma, S. K. (1980) *Practical and Theoretical Consideration involved in Error Analysis*. Indian Journal of Applied Linguistics, 6: 74-83.
43. Sinclair, J. (1982) Reflections on Computer Corpora in English Language Research. In Johansson, S, (1982) ed. *Computer Corpora in English Language Research*. Bergen: Norwegian Computing Centre for the Humanities, pp 1-6.
44. Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University
45. Sinclair, J. (2004) *Developing Linguistic Corpora: a Guide to Good Practice. Corpus and Text — Basic Principles*. Tuscan Word Centre. Available from:
<http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm>
46. Spencer, S. And Edwards, B. (1997) *Text Processing for Modular Awards Series*. Great Britain: Heinemann Educational Publishers
47. Stenson, N. (1978) Induced Errors. In J. Schumann & N. Stenson, (Eds.), *New frontiers in second language learning*. Rowley, Massachusetts: Newbury House Publishing, Inc.
48. Stirling, J. (2011) *Teaching Spelling to English Language Learners*. Raleigh, USA: Lulu

49. Strevens, P. (1968) *Two Ways Looking at Error Analysis*. ERIC: 037 714, 1969. Styles of Learning among American Indians: An Outline for Research. Washington, DC: Center for Applied Linguistics
50. Tarone, E. et al. (1976) *Systematicity/ Variability and Stability/ Instability in Interlanguage System. Papers in Second Language Acquisition: Proceedings of the 6th Annual Conference on Applied Linguistics*. University of Michigan, Language Learning Special Issue.
51. Taubert, W., Chermakova, A. (2004) *Corpus Linguistics. A Short Introduction*. United Kingdom: Continuum
52. ThambiJose.S. F. (2014) *Orthographic Errors Committed by Sophomore Students: A Linguistic Analysis*. Rome: MCSER. Available from:
<http://www.mcser.org/journal/index.php/mjss/article/viewFile/4805/4660>
53. Touchie H.Y. (1986) Second Language Learning Errors Their Types, Causes and Treatment. *JATL Journal*, 8 (1). Available from: http://jalt-publications.org/files/pdf-article/art5_8.pdf
54. Vicedo, J. L. (2004) *Advances in Natural Language Processing: 4th International Conference*, Germany: Springer
55. Wachwicz, K. A., Scott, B. (1999) Software That Listens: It's not a Question of Whether, It's a Question How?. *CALICO Journal*, 16 (3): 253-276.
56. Watson, B. W. and Wood, D. (2002) *Implementation and Application of Automata: 6th International Conference*, Germany: Springer
57. Widdowson, H. G. (2000) On the limitations of linguistics applied. *Applied Linguistics* 21(1): 3-25.

Internet sources

1. Croft , K. (1980) *Contrastive Analysis, Error Analysis and Interlanguage: Three Phases of One Goal* in Readings on English as a Second language for Teachers and Teacher Trainers, Second Edition. Available from
https://www.academia.edu/3760179/Contrastive_Analysis_Error_Analysis_and_Interlanguage_Three_Phases_of_One_Goal [Accessed on March 20, 2015]
2. Lexbook, *Meaning of orthographical in the lexbook online dictionary*. Available from
<http://lexbook.net/en/orthographical> [Accessed on April 2, 2015]

3. Brians, P. (2013) *Hyphenation in Common Errors in English Usage*, Third Edition. Available from <http://public.wsu.edu/~brians/errors/hyphenation.html> [Accessed on March 17, 2015]
4. Donovan, M. *Grammar Rules: Capitalization*. Available from <http://www.writingforward.com/grammar/grammar-rules/grammar-rules-capitalization> [Accessed on April 5, 2015]
5. Byrd, P. *Lecture 6: Commas, Comma Splices, & Fragments* from Weekly Lectures for AL 8460 English Grammar for ESL/EFL Teachers. Department of Applied Linguistics & ESL, Georgia State University, Atlanta, Georgia, USA. Available from http://www2.gsu.edu/~eslhp/grammar/lecture_6/commas.html [Accessed on April 5, 2015]
6. *Definition of carrier in English*. Oxford Dictionaries, Oxford University Press. Available from <http://www.oxforddictionaries.com/definition/english/carrier> [Accessed on April 6, 2015]
7. *Definition of convey in English*. Oxford Dictionaries, Oxford University Press. Available from http://www.oxforddictionaries.com/definition/english/convey#convey__3 [Accessed on May 2, 2015]
8. Sharon, *Wether, Weather, Whether*. Daily Writing Tips. Available from <http://www.dailywritingtips.com/wether-weather-whether/> [Accessed on May 2, 2015]
9. *Common misspellings*. Oxford Dictionaries, Oxford University Press. Available from <http://www.oxforddictionaries.com/us/words/common-misspellings-american> [Accessed on April 20, 2015]
10. Carter, R., McCarthy, M., Mark, G., O’Keeffe, A. (2011) *English Grammar Today Workbook*. Available from <http://dictionary.cambridge.org/grammar/british-grammar/sometimes-or-sometime> [Accessed on May 2, 2015]
11. *Program*. Oxford Dictionaries, Oxford University Press. Available from <http://www.oxforddictionaries.com/definition/learner/program> [Accessed on April 20, 2015]
12. *‘Cannot’ or ‘can not’?* Oxford Dictionaries, Oxford University Press. Available from <http://www.oxforddictionaries.com/words/cannot-or-can-not> [Accessed on March 17, 2015]

Dokumentārā lapa

Bakalaura darbs „Orthographic Error Frequency in Students’ Written Business English” (Ortogrāfisko kļūdu biežums izglītojāmo lietišķajā angļu rakstu valodā) izstrādāts LU Humanitāro zinātņu fakultātē.

Ar savu parakstu apliecinu, ka pētījums veikts patstāvīgi, izmantoti tikai tajā norādītie informācijas avoti un iesniegtā darba elektroniskā kopija atbilst izdrukai.

Autors: Arīna Mihailova

25.05.2015.

Rekomendēju/nerekomendēju darbu aizstāvēšanai

Vadītāja: Nataļja Cigankova

25.05.2015

Recenzents:

Studiju metodiķe:

Darbs iesniegts Anglistikas nodaļā

Darbu pieņēma:

Darbs aizstāvēts bakalaura gala pārbaudījuma komisijas sēdē

2015. gada..... jūnijā, prot. Nr., vērtējums

Komisijas sekretāre: